算法治理对数字平台生态系统的重构效应

——基于微博热搜榜的准自然实验

摘　要： 算法推荐系统作为数字平台的核心信息配置机制，在优化用户体验的同时也引发了信息茧房、注意力垄断等公共治理难题。如何在释放算法价值的同时有效规制其负外部性，成为全球数字治理的核心议题。本文利用 2024 年 11 月中国清朗行动算法治理政策这一准自然实验，采用断点回归时间序列与双重差分相结合的识别策略，基于微博热搜榜高频数据系统评估算法治理对平台内容生态的重构效应。研究发现，算法治理有效推动了注意力分布的去中心化，瓦解了头部话题的垄断格局；分级分类治理导致社会类内容质量显著提升而娱乐类保持稳定，形成差异化的质量改善路径。然而，在注意力总量守恒的零和约束下，监管对高风险领域的强化治理引发了注意力向低风险领域的系统性转移，呈现典型的“水床效应”。在密度层面，社会类呈“少而精”调整，娱乐类呈“多且长”扩张。本研究揭示了算法治理的复杂效应：多样性提升与质量改善的积极成效，与注意力再分配的非预期后果并存，形成“监管初衷”与“生态演化”之间的张力。研究为理解算法治理的系统性后果提供了新的实证证据，对完善分级分类治理体系、探索“激励优质内容 + 动态监测调整 + 跨平台协同治理”的系统性框架具有重要政策启示。

关键词：算法治理；平台生态系统；注意力经济；断点回归时间序列；水床效应
JEL 分类号：D83；L51；L86

Abstract: As the core information allocation mechanism of digital platforms, algorithmic recommendation systems optimize user experience while also triggering public governance challenges such as information cocoons and attention monopolization. How to effectively regulate the negative externalities of algorithms while unleashing their value has become a central issue in global digital governance. This study leverages the quasi-natural experiment of China's Qinglang Action algorithm governance policy implemented in November 2024, employing an identification strategy combining interrupted time series (ITS) with difference-in-differences (DID). Based on high-frequency data from Weibo's trending list, we systematically evaluate the restructuring effects of algorithm governance on platform content ecosystems.

We find that algorithm governance effectively promotes the decentralization of attention distribution, dismantling the monopoly pattern of head topics. Risk-based tiered governance leads to significant quality improvement in social content while entertainment content remains

stable, forming differentiated quality improvement pathways. However, under zero-sum attention constraints where total attention is conserved, intensified regulation of high-risk domains triggers systematic attention transfer toward low-risk domains, exhibiting a pronounced "waterbed effect." At the density level, social content shows a "fewer but refined" adjustment, while entertainment content shows "more and prolonged" expansion. This study reveals the complex effects of algorithm governance: positive outcomes of diversity enhancement and quality improvement coexist with unintended consequences of attention reallocation, creating tension between "regulatory intent" and "ecosystem evolution." The research provides new empirical evidence for understanding the systemic consequences of algorithm governance and offers important policy implications for improving tiered governance systems and exploring a systematic framework of "incentivizing quality content + dynamic monitoring and adjustment + cross-platform coordinated governance."

Keywords: Algorithm governance; Platform ecosystems; Attention economy; Interrupted time series; Waterbed effect
JEL Classification: D83; L51; L86

一、引言

数字平台已成为信息传播、社会互动和经济活动的核心基础设施。作为平台生态系统的“神经中枢”(Jacobides et al., 2018)，算法推荐系统在优化用户体验、提升商业价值的同时，也在深刻塑造着公共信息环境。当算法推荐逐渐成为亿万用户获取信息的“把关人”(Gorwa et al., 2020)，其带来的信息茧房、注意力垄断、内容低质化等问题日益凸显。平台算法通过个性化推荐强化用户既有偏好，可能导致“过滤气泡”效应 (Pariser, 2011)；头部内容凭借算法加权获得不成比例的曝光，形成“赢者通吃”的马太效应 (Hindman, 2018)；为追求点击率最大化，诱导性标题（clickbait）泛滥，信息质量持续下滑 (Chen et al., 2015)。如何有效治理算法推荐系统，已成为全球数字治理的核心议题。

面对算法推荐带来的系统性挑战，各国监管机构纷纷探索治理路径。欧盟《数字服务法》（DSA）要求大型平台披露算法推荐逻辑并提供“无算法”选项；美国学界和政策界围绕《算法问责法案》展开激烈讨论；中国在算法治理领域走在世界前列，《互联网信息服务算法推荐管理规定》（2022 年）和《生成式人工智能服务管理暂行办法》（2023 年）构建了全球最为系统的算法监管框架 (Creemers, 2022)。2024 年 11 月 12 日，国家网信办启动新一轮清朗行动，聚焦算法推荐治理，发布《算法推荐服务专项治理清单指引》，明确 27 项核验标准，要求各大平台“不得利用算法操纵热点话题”、“保障信息内容多样性”、“降低低俗内容权重”，并强化分级分类管理，对社会类等高风险内容实施更严格的审核机制。这一政策冲击为学术界提供了难得的准自然实验场景，引发了本文关注的核心现实问题：当外部监管力量介入平台算法系统，内容生态将如何响应？监管目标能否实现？是否会产生非预期后果？

现有文献在理论层面存在两个关键缺口，制约了对上述问题的系统性回答。第一，缺乏统一的理论框架解释监管引发的多维效应。尽管平台治理研究已积累丰富成果——涵盖平台生态系统理论 (Tiwana et al., 2010)、平台边界资源管理 (Ghazawneh and Henfridsson, 2013)、多边市场治理 (Huber et al., 2013; Parker and Van Alstyne, 2018) 等多个分支——但既有研究多聚焦于单一治理目标（如内容质量或多样性），缺乏将多元治理效应纳入统一分析框架的理论尝试。特别是，算法治理如何同时影响内容质量、注意力分布和生态结构，这些效应之间存在何种内在关联，现有研究尚未提供系统性解释。第二，忽视了监管的非预期后果与目标冲突。注意力是数字平台最核心的稀缺资源 (Davenport and Beck, 2001; Webster, 2016)，用户的总注意力预算在短期内近似固定，形成典型的零和博弈格局。在此约束下，对某一内容类型的监管强化可能引发注意力在不同类型间的系统性再分配。然而，现有算法治理文献 (Gritsenko and Wood, 2022; Ulbricht and Yeung, 2022) 多停留在规范性讨论，缺乏对监管溢出效应的理论刻画。这引出了本文关注的核心理论问题：在资源约束条件下，对高风险内容的监管强化是否会触发平台的资源再配置行为，进而产生“按下葫芦浮起瓢”的水床效应？实证层面同样面临挑战。由于算法系统的“黑箱”特性和高频政策调整，学界难以获得清晰的政策断点和高质量数据，导致算法治理效果的因果识别研究极为稀缺 (Jhaver et al., 2019)。

本文聚焦 2024 年 11 月清朗行动算法推荐治理这一准自然实验，以微博热搜榜为研究对象。在数据层面，本文构建了覆盖政策实施前后的高频面板数据集，为因果识别提

供了充分的时间窗口。在方法层面，采用断点回归时间序列（ITS）识别主效应，辅以双重差分（DID）进行稳健性检验。在理论层面，本文构建了基于资源依赖理论的统一分析框架，将用户注意力视为平台依赖的核心稀缺资源，系统分析算法治理如何通过改变不同内容类型的资源获取成本，触发平台的资源再配置行为。研究发现，算法治理引发了内容生态的多维重构：内容质量呈现分化趋势，高风险内容质量显著提升而低风险内容保持稳定；注意力分布趋于去中心化，头部垄断格局被打破；在注意力总量守恒的零和约束下，社会类内容份额收缩、娱乐类份额扩张，呈现“水床效应”；不同类型内容的调整路径分化，社会类“少而精”、娱乐类“多且长”。

本文的贡献主要体现在三个方面。第一，理论贡献在于构建了基于资源依赖理论的算法治理分析框架。既有研究多聚焦于单一治理目标，缺乏将多元效应纳入统一框架的尝试。本文将注意力作为核心稀缺资源，从资源获取成本与约束条件的双重视角解释监管如何触发平台的系统性响应，提出“水床效应”机制揭示了监管在注意力总量守恒约束下的间接传导路径，拓展了平台治理的理论边界。第二，实证贡献在于利用准自然实验提供了算法治理因果效应的可信估计。由于算法系统的“黑箱”特性和数据可得性限制，算法治理的因果识别研究极为稀缺，本文通过清晰的政策断点和高频数据为政策评估文献增添了中国证据。第三，政策贡献在于揭示了算法治理中多元目标的潜在张力。研究发现，质量提升与多样性改善的同时，可能伴随公共议题曝光度的结构性下降，提示监管部门需要在治理框架中纳入“跨类别溢出效应”的系统考量，建立更加精细化的分类治理策略。

本文的结构安排如下：第二部分梳理相关文献并构建理论框架，推导研究假说；第三部分介绍研究设计和数据来源；第四部分报告实证结果及稳健性检验；第五部分讨论研究发现的理论含义和政策启示。

二、文献回顾与研究假设

（一）文献回顾

1. 平台治理与算法监管

数字平台治理研究经历了从技术视角向制度视角的范式转换。Kango (2025) 在《人本 AI 手册》中系统界定了算法治理的核心概念，强调算法不仅是技术工具，更是影响社会资源分配与权力结构的治理机制。早期研究聚焦于平台作为“双边市场”的经济属性 (Armstrong, 2006; Rochet and Tirole, 2003)，强调网络效应、定价策略与市场竞争。随着平台规模的扩大与社会影响的深化，学者开始关注平台的治理维度。Tiwana et al. (2010) 提出平台生态系统理论，将平台视为由核心架构与互补者构成的共生系统，治理的核心在于协调多方利益、维护生态健康。Ghazawneh and Henfridsson (2013) 进一步提出边界资源概念，分析平台如何通过 API、规则与激励机制管理开发者行为。Parker and Van Alstyne (2018) 则从网络效应视角系统阐述了平台治理的独特逻辑，指出平台需要同时管理供给端（内容生产者）与需求端（用户），任何治理干预都可能产生跨边溢出效应。Jacobides et al. (2018) 从生态系统视角分析了平台企业的治理逻辑与价值创造范式，Huber et al. (2013) 进一步系统梳理了平台监管中契约治理与关系治理的互补与替代机制。焦豪 (2023) 提出“数字平台生态观”这一理论新视角，强调数字经济时代企业需

要建立与自身资源相匹配的数字平台，通过与利益相关者共同创造价值形成持续竞争优势。Degen and Gleiss (2025) 则从监管设计视角提出 “平台治理标准类型” 分类框架，主张针对不同类型平台实施差异化监管策略，为本文的分级分类治理分析提供了理论参照。

算法推荐系统的治理构成平台治理的新兴分支。算法作为平台的 “隐性架构”(Gillespie, 2014)，通过个性化推荐深度介入用户的信息获取过程，成为事实上的 “信息把关人”(Gorwa et al., 2020)。Bucher (2018) 指出，算法权力体现在其 “可供性”（affordance），算法不仅反映用户偏好，更主动塑造用户行为。这一视角在后续研究中得到进一步拓展，Gillespie (2018) 系统分析了平台作为 “互联网守门人” 的角色演变，强调算法已成为重构信息流通逻辑的核心力量；Katyal (2019) 则从法律视角分析了人工智能时代私人问责的边界问题。这一双向互动使得算法治理面临独特挑战：监管者既要约束算法的负面效应，又要避免过度干预导致系统效率损失 (Yeung, 2018)。

方法论层面，算法审计（algorithm audit）为理解算法行为提供了实证工具。Sandvig et al. (2014) 系统梳理了算法审计的研究方法，提出了检测互联网平台歧视行为的实证框架；Diakopoulos (2016) 进一步将算法问责的理念应用于新闻推荐领域，发现算法推荐存在显著的议题设置效应。近年来，算法审计方法得到快速发展：Mousavi et al. (2024) 以 TikTok 为对象审计了平台提供的算法解释功能，揭示了算法透明性承诺与实际操作之间的差距；Wang et al. (2024) 对比分析了 Twitter/X 算法推荐与时间线排序对新闻内容质量的差异化影响，发现算法推荐虽然降低了信息数量但提升了内容质量；Hosseinmardi et al. (2024) 采用 “反事实机器人” 方法因果估计了 YouTube 推荐算法的效应，为算法审计的因果识别提供了方法论创新。国内学者亦对算法审计方法进行了本土化探索，师文和陈昌凤 (2023) 基于计算实验方法系统审计了国内主流平台算法的 “主流化” 偏向与 “个性化” 特质，为理解中国情境下的算法行为提供了实证基础。这些研究为本文的实证设计提供了方法论参考。

全球范围内，算法治理呈现差异化路径。欧盟采取 “透明度 + 问责” 模式，《通用数据保护条例》（GDPR）赋予用户 “解释权”，《数字服务法》（DSA）要求平台披露推荐逻辑并接受第三方审计。Panigutti et al. (2025) 系统梳理了 DSA 框架下算法风险调查的方法论路径，为平台合规与外部监督提供了操作指南。美国延续 “自律为主” 传统，但围绕《算法问责法案》的讨论日趋激烈 (Katyal, 2019)。李三希等 (2023) 从产业组织视角考察了市场竞争对企业数据行为的影响，发现竞争能够有效约束企业过度收集消费者信息的动机，这为理解数据治理的市场化路径提供了理论基础。中国构建了全球最为系统的算法监管框架：2021 年《互联网信息服务算法推荐管理规定》确立分级分类原则，2023 年《生成式人工智能服务管理暂行办法》将监管延伸至大模型领域 (Creemers, 2022)。Xu (2024) 深入分析了中国算法监管的 “打开黑箱” 策略，指出中国监管模式强调 “风险分级 + 事前审核” 的制度组合；Sheehan (2023) 进一步揭示了中国 AI 监管的政策制定过程与多部门协调机制。Ulbricht and Yeung (2022) 指出，算法监管已成为全球治理的成熟议题，各国在透明度、问责与风险分级等维度形成了差异化的制度路径。2024 年 “清朗行动” 算法专项治理则以 27 项核验标准推动规则落地，体现了 “风险分级 + 协同治理” 的制度逻辑 (Gritsenko and Wood, 2022)。这一制度演进为本文提供了清晰的政策断点。

2. 注意力经济与信息分配

注意力经济理论为理解平台内容生态提供了资源视角。Simon (1996) 最早指出，在信息过载时代，稀缺的不是信息本身而是处理信息的注意力。Davenport and Beck (2001) 将注意力定义为 “对特定信息的聚焦心智参与”，强调其稀缺性、竞争性与货币化潜力。Webster (2016) 进一步提出 “注意力市场” 概念，将媒体竞争重新框定为对用户时间与心智份额的争夺。Wu (2017) 从历史视角系统梳理了注意力商业化的演进历程，揭示了数字时代注意力稀缺与算法分配的内在逻辑。O’Reilly et al. (2024) 提出 “算法注意力租金” 概念，指出数字平台通过算法控制实现对用户注意力的系统性提取，形成独特的市场权力来源；Heitmayer (2025) 进一步发展了 “注意力经济第二波” 理论，将注意力概念化为社交媒体时代的 “通用符号货币”。喻国明和刘彧晗 (2023) 基于元传播视角提出，数字平台竞争已从 “信息竞争” 转向 “认知竞争”，这一范式转型深刻重塑了注意力资源的获取与分配机制。

在数字平台情境下，注意力的分配呈现高度不均衡特征。Hindman (2018) 通过大规模网络流量分析揭示了 “赢者通吃” 格局，发现少数头部内容获取了不成比例的用户注意力，形成典型的幂律分布。Cui and Kertész (2023) 以中国微博为研究对象，发现热门话题的注意力竞争呈现显著的 “富者愈富” 效应，平台干预能够有效调节这一格局。算法推荐系统可能加剧这一趋势：通过放大用户既有偏好，算法创造了 “过滤气泡”（filter bubble）与 “回音室”（echo chamber）效应 (Pariser, 2011; Sunstein, 2018)。然而，也有研究对此提出质疑。Guess et al. (2023) 基于大规模实验发现，关闭算法推荐并未显著改善用户的信息多样性，暗示算法效应可能被高估。Hosanagar et al. (2014) 的研究也表明，推荐系统对消费者信息多样性的影响存在高度异质性。Ahmmad et al. (2025) 对过滤气泡与回音室研究进行了系统综述，指出算法对青年群体的影响存在显著异质性，过度简化的 “信息茧房” 叙事可能掩盖了更复杂的现实。喻国明和刘彧晗 (2024) 进一步指出，“信息茧房” 议题可能被过度渲染，个性化推荐并不必然导致用户视野收窄，算法推荐的实际效应需要更多实证检验。这一争论凸显了算法治理效果评估的重要性与复杂性。

注意力的零和特性对治理具有重要含义。Wu (2017) 指出，平台经济的核心商业模式是 “注意力商人”，即通过免费内容吸引用户注意力，再将注意力出售给广告商。在此模式下，用户的总注意力构成平台收入的 “天花板”。这意味着，对某一类型内容的监管干预可能引发注意力在不同类型间的再分配，本文将这一机制概念化为 “水床效应”。

3. 内容质量与信息诱导

内容质量劣化是算法治理的核心关切之一。信息差理论（information gap theory）为理解诱导性内容提供了心理学基础。Loewenstein (1994) 指出，当个体意识到知识缺口时会产生强烈好奇心，驱使其采取行动填补缺口。诱导性标题（clickbait）正是利用这一心理机制，通过制造标题与内容之间的语义悬念来操纵用户的注意力分配 (Golman and Loewenstein, 2018)。Chakraborty et al. (2016) 通过机器学习方法识别 clickbait 特征，发现疑问句式、悬念词汇、情感激发是典型策略。Chen et al. (2015) 分析了 clickbait 的经济逻辑，指出在注意力竞争中，夸张标题能够获得更高点击率，但会损害用户信任与长期价值。

内容质量治理面临 “逐底竞争” 困境。Munger (2020) 指出，在算法驱动的信息环境

中，内容生产者面临集体行动困境：个体理性（追求点击率）与集体理性（维护信息质量）相冲突，导致低质内容驱逐高质内容的“柠檬市场”效应。Berman and Katona (2020) 系统分析了算法策展对内容质量的影响，发现内容质量呈现分化趋势，专业媒体坚守内容标准，而用户生成内容更易陷入流量导向。Jhaver et al. (2019) 基于 Reddit 的实证研究发现，内容审核透明性与用户信任之间存在显著正相关，这为治理干预提供了行为依据。平台治理的干预逻辑在于改变激励结构，使高质量内容获得更高回报 (Gillespie, 2018)。值得注意的是，内容审核可能产生非预期的溢出效应。Cima et al. (2025) 采用双重差分方法研究大规模内容审核干预的异质性效应，发现审核行为在不同用户群体间产生差异化影响，部分群体甚至出现“反弹”现象；Gomes and Sultan (2024) 则从公共健康视角分析了内容审核对数字减害实践的意外阻碍。这些发现提示，算法治理的效果评估需要关注监管的间接传导路径与非预期后果。然而，既有研究多聚焦于平台自治，对外部监管如何影响内容质量的实证分析相对匮乏。

4. 现有研究的不足与本文定位

综上所述，现有文献在三个方面存在不足，构成本文的理论切入点。

第一，缺乏统一框架整合多维效应。平台治理文献分散于内容质量、信息多样性、用户行为等子领域，缺乏将多元治理效应纳入统一分析框架的理论尝试。尽管学者们在平台治理 (Parker and Van Alstyne, 2018; Tiwana et al., 2010)、算法权力 (Bucher, 2018; Katyal, 2019) 等领域取得了重要进展，但鲜有研究将这些分散的理论视角系统整合。本文引入资源依赖理论，将用户注意力视为平台依赖的核心稀缺资源，从资源再配置视角统一解释内容质量分化、注意力去中心化、水床效应等多维效应。

第二，忽视监管的非预期后果。既有算法治理文献多聚焦于规范性讨论或单一目标评估 (Gritsenko and Wood, 2022; Ulbricht and Yeung, 2022)，缺乏对监管溢出效应的系统分析。本文提出“水床效应”概念，刻画在注意力零和约束下监管干预的间接传导路径，揭示质量提升与公共议题曝光度下降之间的潜在张力。

第三，因果识别研究稀缺。由于算法系统的“黑箱”特性和数据获取难题，既有研究多为描述性分析或规范性讨论。虽然 Sandvig et al. (2014) 和 Diakopoulos (2016) 等学者发展了算法审计的方法论，但这些研究主要聚焦于算法行为的描述性刻画，而非政策干预的因果效应评估。本文利用 2024 年清朗行动提供的准自然实验场景，采用 ITS 与 DID 相结合的因果识别策略，提供算法治理效果的可信估计。

（二）理论框架与研究假设

1. 资源依赖视角下的算法治理分析

资源依赖理论（Resource Dependence Theory）认为，组织的生存与发展取决于其从外部环境获取关键资源的能力 (Pfeffer and Salancik, 1978)。当组织高度依赖某一稀缺资源时，该资源的供给者将对组织行为产生深远影响。Yu and Sekiguchi (2024) 系统梳理了平台依赖型创业的研究文献，指出平台生态系统中的资源依赖关系具有独特的双向性与动态性特征。本文将这一理论框架应用于数字平台情境，提出核心命题：用户注意力构成了数字平台最核心的稀缺资源，算法治理的本质是改变平台获取和配置注意力资源的成本结构与约束条件。

在数字平台生态系统中，注意力资源具有三个决定性特征 (Berman and Katona, 2020; Oestreicher-Singer and Sundararajan, 2012)。第一，稀缺性：用户日均使用时长存在生理和时间约束，短期内缺乏弹性。第二，竞争性：注意力分配给某一话题必然减少其他话题的份额，形成典型的零和博弈格局。第三，不可再生性：注意力一旦消耗无法恢复。这三个特征共同构成了本文分析的基础约束——注意力总量守恒：各类内容获得的注意力之和等于平台可获取的总注意力预算。

基于资源依赖理论，算法治理通过两条路径影响平台行为：**成本路径**与**约束路径**。

成本路径：《算法专项治理清单指引》通过分级分类监管，对不同风险类别的内容施加差异化的合规要求，直接改变了平台的资源获取成本。监管后各类内容的成本变化取决于其风险等级和监管强度，且成本变化与风险等级正相关。由于社会类内容涉及公共利益、舆情风险等敏感领域，其风险等级显著高于娱乐类，导致社会类内容的合规成本大幅上升，而娱乐类成本基本稳定。这一差异化成本冲击将驱动平台对高风险内容实施更严格的质量筛选（对应假设 H1），并在注意力总量守恒约束下引发跨类别的资源再分配（对应假设 H3、H4）。

约束路径：监管同时施加了多样性约束，要求平台 “不得利用算法操纵热点话题”、“保障信息内容多样性”。这直接改变了平台的可行域结构，限制了头部话题的注意力垄断能力（对应假设 H2）。

面对差异化的成本冲击与新增约束，平台作为理性主体将重新优化其资源配置策略。平台的目标是在注意力总量守恒和多样性约束条件下最大化净收益。最优化条件要求各类内容的边际净收益相等。当社会类的合规成本上升时，平台需要减少社会类的注意力配置以提高其边际收益，并将释放的资源重新配置至成本较低的类别。在注意力总量守恒约束下，这一再配置过程产生了本文提出的 “水床效应”：社会类注意力下降必然导致娱乐类和其他类别注意力上升。

基于上述理论框架，本文提出四项研究假设。这四项假设共同构成资源依赖视角下算法治理效应的完整图景：H1 刻画成本路径的直接效应（质量筛选），H2 刻画约束路径的直接效应（多样性提升），H3 刻画资源总量约束下的再配置效应（水床效应），H4 刻画再配置效应在微观层面的具体表现（密度分化）。

2. 假设 H1：内容质量分化效应

理论逻辑：资源依赖理论强调，组织会根据资源获取成本调整其行为策略 (Pfeffer and Salancik, 1978)。分级分类监管的核心在于根据内容的舆论风险等级实施差异化监管强度。《算法专项治理清单指引》对涉及公共事件、舆情风险的社会类内容提出了更高的信息来源、事实核验和价值导向标准，而对娱乐类内容多采取 “守住底线” 的防御性姿态。

从资源依赖视角看，这一差异化监管直接转化为平台的差异化合规成本。社会类内容的合规成本大幅上升，包括强化内容审核成本、信息来源核查成本和违规问责风险成本；娱乐类内容的合规成本基本稳定。根据前述理论模型，当社会类合规成本上升时，平台的最优响应是提高社会类内容的质量门槛，通过 “事前筛选” 淘汰低质内容，从而降低单位内容的合规风险。换言之，社会类的质量门槛与其合规成本正相关，而娱乐类质量门槛基本不变。

在社会类内部，机构媒体相对自媒体将展现出更显著的质量提升。这源于两个机制：第一，源头筛选机制，平台在高风险领域优先选择机构媒体作为信息源头，因为机构媒体的内容规范性和可信度更高 (He et al., 2022)，能够有效降低平台的合规风险；第二，信号显示机制，平台通过提升机构媒体占比向监管者传递合规信号，以获取监管合法性。

基于上述分析，提出假设 H1：

H1（内容质量分化）：算法治理导致不同风险类别内容的质量呈现分化趋势，高风险内容质量显著提升，低风险内容质量基本稳定。

H1a：社会类话题的信息差指数（衡量标题信息诱导程度）在监管后显著下降

H1b：社会类内部，机构媒体相对自媒体的信息差净下降（DID 估计量显著为负）

H1c：娱乐类话题的质量指标在监管前后无显著变化

3. 假设 H2：注意力去中心化效应

理论逻辑：资源依赖理论指出，外部约束的变化将迫使组织调整其资源配置模式 (Pfeffer and Salancik, 1978)。《算法专项治理清单指引》明确要求平台 “不得利用算法操纵热点话题”、“保障信息内容多样性”。这些监管条款构成了平台资源配置的硬约束边界，直接限制了头部话题的注意力垄断能力。

从资源配置视角看，多样性约束改变了平台的可行域结构。监管要求政策实施后话题分布的多样性水平（以 Shannon 熵衡量）必须高于政策实施前的最低标准，即监管提高了平台内容多样性的下限要求。该约束迫使平台压缩头部话题的资源配额、提升长尾话题的曝光机会。在注意力总量守恒的前提下，这意味着原本高度集中于少数话题的注意力将被重新分配至更广泛的话题集合。平台通过限制同一话题 ID 的上榜频次、设置话题轮换机制，从供给端切断了头部话题的垄断路径。

同时，监管还通过改变平台的风险收益函数，激励其主动提升内容多样性。单一话题过度曝光的违规概率在监管后显著上升，平台因话题过度集中而遭受监管处罚的预期成本相应增加。平台为规避监管风险而主动调低头部话题权重、提升长尾话题推荐权重。

基于上述分析，提出假设 H2：

H2（注意力去中心化）：算法治理打破了热搜榜单的赢者通吃格局，注意力从少数头部话题向广泛的长尾话题再分配。

H2a：Shannon 熵在监管后显著上升

H2b：HHI 集中度指数在监管后显著下降

H2c：单条话题重复上榜次数和连续在榜时长在监管后显著下降

4. 假设 H3：注意力再分配效应（水床效应）

理论逻辑：水床效应是资源依赖理论在注意力总量守恒约束下的核心推论。资源依赖理论强调，当某一资源获取渠道受阻时，组织会转向替代性资源渠道 (Pfeffer and Salancik, 1978)。在本文情境下，监管提高了社会类内容的获取成本，平台作为理性主体将调整其资源配置，将注意力从高成本领域转向低成本领域。

这一再分配过程同时受到供给侧和需求侧的双重驱动。从供给侧看，监管实施后社会类的 “有效供给” 因质量门槛提高和合规成本上升而下降。根据前述理论模型，平台选择的最优社会类内容数量与其合规成本呈负相关关系：当社会类合规成本上升时，平

台倾向于减少社会类内容的供给数量。从需求侧看，用户的总注意力需求在短期内保持稳定，不会因监管而显著改变。在供需缺口下，用户将注意力转向替代性内容。由于娱乐类内容的审核成本相对较低，其成为承接溢出注意力的“蓄水池”。

这一“成本驱动”的再配置逻辑与用户行为的被动调整相互强化，共同塑造了“社会类紧缩与娱乐类扩张”的结构性格局。水床效应的存在可通过验证注意力总量守恒来间接证实：如果用户的总注意力预算在短期内保持稳定，则算法治理不应显著改变平台的总注意力供给量，而仅改变其在不同类型间的分配结构。具体表现为：平台总注意力保持基本稳定，但社会类注意力份额下降，娱乐类注意力份额相应上升。

基于上述分析，提出假设 H3：

H3（注意力再分配/水床效应）：在注意力零和约束下，对社会类的监管强化将引发注意力向娱乐类的系统性转移。

H3a：社会类话题在热搜榜中的份额在监管后显著下降

H3b：娱乐类话题在热搜榜中的份额在监管后显著上升

H3c：周度总热度在监管前后无显著变化，验证注意力零和约束

5. 假设 H4：注意力密度分化效应

理论逻辑：注意力密度分化是水床效应在微观层面的具体表现，揭示了平台在宏观配额调整基础上的精细化响应策略。资源依赖理论指出，组织在资源约束下会采取差异化策略以最大化资源利用效率 (Pfeffer and Salancik, 1978)。在本文情境下，社会类与娱乐类因其风险特征不同，将呈现截然不同的调整路径。

社会类预期呈现“少而精”的调整路径。根据前述分析，监管显著提高了社会类内容的准入门槛，低质内容在更严格的审核标准下被淘汰出局，导致上榜话题数量明显减少。与此同时，通过质量筛选得以留存的高质量话题因其信息价值和公共属性获得平台更长时间的曝光支持。这一策略符合资源依赖理论的核心逻辑：在资源获取成本上升时，组织会更加谨慎地使用资源，确保每单位资源都能产生更高价值。在话题数量减少而总在榜时长相对稳定的综合作用下，社会类的单条话题平均在榜时长呈现上升态势，即注意力密度提高。

娱乐类预期呈现“多且长”的扩张格局。在注意力总量守恒的约束下，社会类受挤压释放的配额需要寻找新的“蓄水池”，审核成本相对较低的娱乐类内容成为承接溢出注意力的天然选择。平台为填补社会类紧缩造成的配额缺口，主动增加娱乐类话题的上榜数量，并延长单条话题的在榜时长，形成话题数量和单条时长同步扩张的局面。

基于上述分析，提出假设 H4：

H4（注意力密度分化）：在注意力再分配过程中，不同类型内容的注意力密度呈现差异化演化路径。

H4a：社会类单条话题的平均在榜时长在监管后保持稳定或上升，体现“少而精”调整

H4b：娱乐类的总在榜时长、时间份额、单条话题平均时长在监管后显著上升，体现“多且长”调整

（三）理论框架总结

综上所述，本文构建了基于资源依赖理论的算法治理分析框架（见图1）。该框架的核心逻辑是：算法治理通过分级分类监管改变了不同内容类型的资源获取成本（成本路径）并施加多样性约束（约束路径），平台作为资源依赖者在新的成本结构与约束条件下重新优化资源配置，最终导致内容生态的多维重构。

四项假设之间的逻辑关系可概括为：H1（内容质量分化）是成本路径的直接效应，高风险内容面临更高合规成本从而提升质量门槛，体现了资源依赖理论中“成本上升导致筛选强化”的基本逻辑；H2（注意力去中心化）是约束路径的直接效应，多样性监管约束迫使平台打破头部垄断格局，体现了“外部约束改变资源配置模式”的理论预测；H3（水床效应）是资源总量约束下的再配置效应，高成本领域紧缩、低成本领域扩张，体现了“资源获取渠道受阻时组织转向替代渠道”的核心命题；H4（密度分化）是资源再配置在微观层面的表现，社会类“少而精”、娱乐类“多且长”，体现了“差异化策略最大化资源利用效率”的组织行为特征。这一理论框架将多元治理效应纳入统一的资源依赖逻辑，为理解算法治理的系统性后果提供了分析工具。

三、样本选择与数据来源

（一）数据来源与样本筛选

本研究使用新浪微博热搜榜的历史数据作为核心数据源。微博作为中国最大的社交媒体平台之一，其热搜榜单通过算法实时聚合用户浏览、搜索、讨论等行为数据，动态生成并展示当前最受关注的 50 个话题，是观察算法治理效应的理想场景。Liu et al. (2025) 揭示了微博热搜的生成机制，指出热搜榜单是用户行为、平台算法与商业利益多方互动的产物，其运作逻辑远比“自然涌现”更为复杂。数据采集的时间窗口为 2024 年 2 月 1 日至 2025 年 10 月 1 日，跨度共计 608 天，完整覆盖了 2024 年 11 月 12 日算法治理政策实施的前后阶段，其中政策前观测期为 253 天，政策后观测期为 355 天。

原始数据包含话题名称、热度值（微博官方计算的综合指标，整合了阅读量、讨论量、搜索量等多维度数据）、上榜时长、话题主持人分类（微博平台预设的 81 个细分类别）、时间戳（精确到分钟级别）等关键维度。为确保数据质量，本文对原始数据实施了三步筛选：第一，剔除时间戳缺失或异常的记录；第二，删除空话题或系统占位符；第三，排除平台未明确分类的“其他”类别话题（该类别含义模糊，无法纳入类别对比分析）。经过筛选，最终样本包含 253,227 条观测值，对应约 215,083 个独立话题，覆盖 608 个日样本。

（二）话题分类方法

本研究直接沿用微博平台预设的主持人分类字段作为话题分类的基础。选择平台原始分类的理由有三：第一，减少测量误差，平台分类基于大规模用户行为数据与算法模型训练，相较于研究者的二次编码具有更高的稳定性；第二，贴近平台决策逻辑，算法治理政策的实施主体是平台本身，使用平台原始分类可最大程度还原算法调整的真实作用路径；第三，提升研究可复制性，平台分类数据可直接获取且公开透明。

虽然原始数据包含 81 个细分类别，但为聚焦理论假设中的核心对比——“高风险社

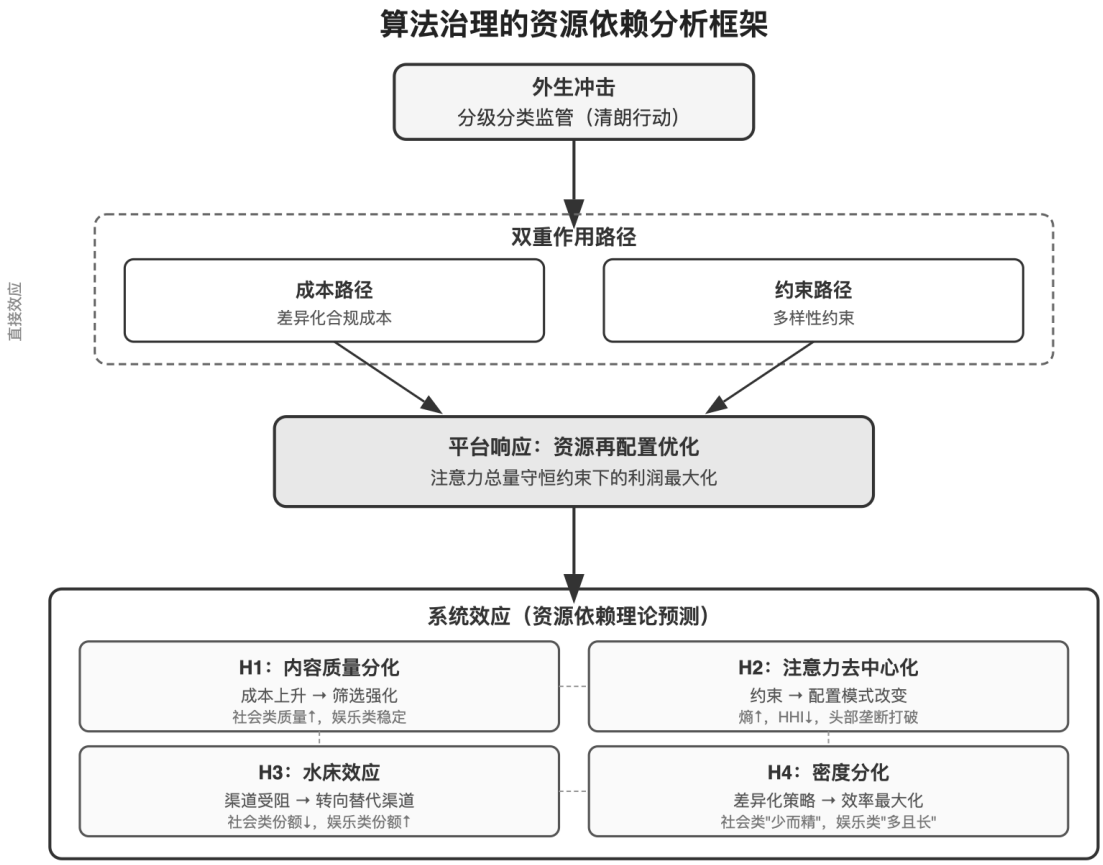
算法治理的资源依赖分析框架
外生冲击
分级分类监管（清朗行动）
双重作用路径
成本路径
差异化合规成本
约束路径
多样性约束
直接效应
平台响应：资源再配置优化
注意力总量守恒约束下的利润最大化
系统效应（资源依赖理论预测）
H1：内容质量分化
成本上升 → 筛选强化
社会类质量↑，娱乐类稳定
H2：注意力去中心化
约束 → 配置模式改变
熵↑，HHI↓，头部垄断打破
H3：水床效应
渠道受阻 → 转向替代渠道
社会类份额↓，娱乐类份额↑
H4：密度分化
差异化策略 → 效率最大化
社会类"少而精"，娱乐类"多且长"

图 1 理论分析框架

会类内容” 与 “低风险娱乐类内容” 之间的差异化效应——本文选取了占比排名前十的类别作为主分析对象。这十大类别合计覆盖约 86% 的样本量，具体分布见表1。

表 1 Top 10 话题类别分布

类别	占比	累计占比	风险等级	核心分析组
社会	34.82%	34.82%	高风险	社会类
明星	11.86%	46.68%	低风险	娱乐类
体育	10.43%	57.11%	中风险	—
时事	6.52%	63.63%	中风险	—
游戏	4.82%	68.45%	低风险	娱乐类
明星-内地	4.61%	73.06%	低风险	娱乐类
综艺	4.30%	77.36%	低风险	娱乐类
电视剧	3.17%	80.53%	低风险	娱乐类
搞笑	2.69%	83.22%	低风险	娱乐类
财经	2.46%	85.68%	中风险	—

基于理论假设中关于 “监管强度的非对称性”，本文将上述类别聚合为两个核心对比组：社会类（占比 34.82%），主要涵盖公共事件、民生议题、社会现象等具有较强舆论属性的内容，根据《算法专项治理清单指引》第 11 条和第 23 条，此类内容被纳入平台重点监测范围，面临更高的信息真实性、来源可溯性要求，构成典型的 “高风险、高合规成本” 类别；娱乐类（包括明星、明星-内地、游戏、综艺、电视剧、搞笑等六个子类别，合计占比 31.45%），主要涵盖明星动态、娱乐八卦、游戏资讯、综艺节目、影视剧集、搞笑内容等泛娱乐领域话题，其特征是用户参与度高但舆论风险相对可控，监管多采取 “守住底线” 的防御性姿态，构成典型的 “低风险、低合规成本” 类别。选择这两个类别作为核心对比组，是因为二者在监管强度、内容属性、用户群体等维度上呈现出最鲜明的两极分化特征，这种 “极端对比” 策略有助于清晰识别算法治理的差异化效应。

（三）变量定义与测量

本研究构建了涵盖多样性、质量、注意力配置、注意力密度四个维度的指标体系，以全面检验算法治理对平台生态系统的重构效应。表2汇总了所有核心变量的定义、测量方法、文献来源及对应假说。

1. 质量指标（H1）

信息差指数基于 Loewenstein (1994) 的信息差理论（information gap theory），该理论指出当个体意识到知识缺口时会产生强烈好奇心，驱使其采取行动填补缺口。诱导性标题（clickbait）正是利用这一心理机制，通过制造标题与内容之间的语义悬念来操纵用户的注意力分配 (Golman and Loewenstein, 2018)。本研究采用基于关键词词频统计的方法构建信息差指数，该方法在社交媒体内容分析中得到广泛应用 (Chakraborty et al., 2016)。

具体而言，本文识别话题标题中制造好奇缺口的三类语言特征：疑问词（为何、为什么、怎么、什么、原因、真相）、悬念词（曝、竟、暗示、去向、结局、神反转、揭秘、内幕、背后）、疑问标点（?、?）。本研究采用关键词词频聚合方法构建信息差指数：通过对标题文本进行分词处理，统计上述三类特征词的出现频次，将标题中出现任一特征词

的情况标记为存在信息差。该方法遵循内容分析领域的标准化编码流程，在保证测量效度的同时兼顾了大规模文本分析的可操作性。该二元测度直接捕捉了标题是否利用 “知识缺口” 策略吸引注意力，值越高表示信息诱导程度越强，内容质量越低。

为检验 H1b 中提出的 “机构媒体相对优势” 假说，本文对话题的信源属性进行了分类。根据 He et al. (2022) 对中国舆论生态的研究，机构媒体在内容规范性和信息可信度方面具有显著优势，构成平台治理中的 “质量基准”。本研究基于人工标注数据（manual_host_labels.csv）对话题主持人进行机构媒体/自媒体分类，该标注结合了启发式规则和人工校验。机构媒体定义为：传统主流媒体（如央视新闻、澎湃新闻、新京报）、政务账号（如公安部、各地卫健委）及新闻门户账号。机构媒体虚拟变量赋值为 1，自媒体及个人账号赋值为 0。

2. 多样性指标（H2）

Shannon 熵是信息论中衡量分布不确定性的经典指标，广泛应用于推荐系统多样性的评估 (Oestreicher-Singer and Sundararajan, 2012)。该指标值越大，表示注意力分布越均匀，头部垄断程度越低。定义为：

$$H(p) = -\sum_{i=1}^{N} p_i \ln p_i \tag{1}$$

其中，p_i 为第 i 个话题获得的注意力份额（以在榜时长占比衡量），N 为观测日内上榜话题总数。当所有话题获得等额注意力时（$p_i = 1/N$），Shannon 熵达到最大值 $\ln N$；当注意力完全集中于单一话题时（$p_1 = 1, p_{i\neq 1} = 0$），熵值降至 0。

HHI 指数源于产业组织理论，用于衡量市场集中度 (Berman and Katona, 2020)。在注意力经济情境下，HHI 反映了少数头部话题对总注意力的垄断程度。定义为：

$$\text{HHI} = \sum_{i=1}^{N} p_i^2 \tag{2}$$

HHI 的取值范围为 $[1/N, 1]$。当注意力完全均匀分布时，$\text{HHI} = 1/N$，集中度最低；当注意力完全集中于单一话题时，$\text{HHI} = 1$，集中度最高。

3. 注意力配置指标（H3）

注意力份额衡量特定类别内容在热搜榜总注意力资源中的占比，反映了平台算法在不同内容类型间的资源配置偏好。Hosanagar et al. (2014) 在研究用户注意力的稀缺性时指出，注意力资源具有零和性质，对某一类型的注意力分配必然减少其他类型的份额。本研究定义注意力份额为：

$$\text{Share}_{it} = \frac{\sum_{j \in i} \text{Duration}_{jt}}{\sum_{k} \text{Duration}_{kt}} \tag{3}$$

其中，i 代表内容类别（社会类或娱乐类），t 代表时间（日度），Duration_{jt} 为话题 j 在 t 日的在榜时长，分子为类别 i 内所有话题的总在榜时长，分母为当日所有话题的总在榜时长。

周度总热度用于验证 H3c 中提出的 “注意力零和约束”。如果用户的总注意力预算

在短期内保持稳定，则算法治理不应显著改变平台的总注意力供给量，而仅改变其在不同类型间的分配结构。定义为：

$$\text{TotalHeat}_w = \sum_{t \in w} \sum_{j} \text{Heat}_{jt} \tag{4}$$

其中，w 代表周（7 天为一个观测窗口），Heat_{jt} 为话题 j 在 t 日的热度值。

4. 注意力密度指标（H4）

注意力密度衡量单位内容获得注意力的强度，揭示了在总量重新分配的同时，单条话题获得注意力的深度变化 (Leskovec et al., 2009)。本研究定义类内注意力密度为：

$$\text{Density}_{it} = \frac{\sum_{j \in i} \text{Duration}_{jt}}{\text{Count}_{it}} \tag{5}$$

其中，Count_{it} 为类别 i 在 t 日的上榜话题数量，分子为类别总在榜时长，分母为话题数量。该指标分解了总注意力变化的两个维度：广度（话题数量）与深度（单条话题强度）。

表 2 变量定义与测量

变量名称	变量符号	变量定义
质量指标（*H1*）		
信息差指数	*InfoGap*	二元变量，标题含疑问词/悬念词/疑问标点=1，否则 =0
机构媒体	*Institutional*	主流媒体/政务账号/新闻门户 =1，自媒体及个人 =0（基于人工标注）
多样性指标（*H2*）		
Shannon 熵	*H*	$H(p) = -\sum_{i=1}^{N} p_i \ln p_i$，$p_i$ 为第i 个话题的注意力份额
HHI 指数	*HHI*	$\text{HHI} = \sum_{i=1}^{N} p_i^2$，衡量注意力集中度
注意力配置指标（*H3*）		
注意力份额	*Share*	类别i 在t 日的在榜时长占全部话题在榜时长的比例
周度总热度	*TotalHeat*	一周内所有话题热度值的加总
注意力密度指标（*H4*）		
单条话题平均时长	*Density*	类别总在榜时长除以该类别上榜话题数量

（四）计量模型设定

本研究采用断点回归时间序列（ITS）作为主要识别策略，辅以双重差分（DID）进行稳健性检验。Anglin et al. (2023) 系统比较了 DID 与比较断点时间序列（CITS）在重复测量设计中的因果推断效度，发现两种方法的组合使用能够有效提升估计的稳健性；Jiang et al. (2024) 进一步梳理了 ITS 设计在政策评估中的应用规范。本文的方法论组合策略正是基于上述方法论文献的建议。ITS 模型用于检验四项核心假设的主效应；DID 模型专门用于 H1（内容质量分化）的稳健性检验，通过对比社会类内部机构媒体与自

媒体的相对变化，强化因果识别。

1. 模型 1：内容质量分化检验（H1）

本文采用两阶段检验策略。

第一阶段（ITS 主效应检验）：通过分组 ITS 回归对比社会类与娱乐类在信息差指数上的政策前后变化：

$$\text{InfoGap}_{it} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{Trend}_t + \beta_3 (\text{Post}_t \times \text{Trend}_t) + \gamma_w + \delta_q + \theta \text{Holiday}_t + \varepsilon_{it} \quad (6)$$

分别对社会类和娱乐类样本运行该模型，检验各自的质量变化趋势。β_1 捕捉政策的即时效应，β_3 捕捉政策的趋势效应。模型控制星期固定效应 γ_w、季度固定效应 δ_q 和节假日效应 θ，以剔除时间序列中的周期性波动。

第二阶段（DID 稳健性检验）：在社会类内部构建 DID 模型，以机构媒体话题为处理组、自媒体话题为对照组：

$$\text{InfoGap}_{it} = \alpha + \delta (\text{Post}_t \times \text{Institutional}_i) + \gamma_i + \lambda_t + \varepsilon_{it} \quad (7)$$

其中，Institutional_i 为机构媒体虚拟变量，γ_i 为话题固定效应，λ_t 为时间固定效应。δ 为 DID 估计量，捕捉政策实施后机构媒体相对自媒体的信息差净变化。DID 识别的有效性依赖于平行趋势假设，本文将在稳健性检验部分报告平行趋势检验结果。

2. 模型 2：注意力去中心化检验（H2）

基于日度时间序列数据，本文采用 ITS 模型检验政策实施后 Shannon 熵和 HHI 的断点变化：

$$Y_t = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{Trend}_t + \beta_3 (\text{Post}_t \times \text{Trend}_t) + \gamma_w + \delta_q + \theta \text{Holiday}_t + \varepsilon_t \quad (8)$$

其中，Y_t 为 t 日的 Shannon 熵或 HHI，Post_t 为政策实施虚拟变量（2024 年 11 月 12 日及之后 =1，之前 =0），Trend_t 为线性时间趋势。为控制时间序列中的周期性波动，模型纳入星期固定效应 γ_w（以周一为基准）和季度固定效应 δ_q（以第一季度为基准），以捕捉工作日与周末、不同季节之间的系统性差异。此外，模型引入节假日虚拟变量 Holiday_t 控制中国法定节假日（包括春节、国庆、清明、端午、中秋等）对热搜话题分布的冲击效应。标准误采用异方差稳健标准误（HC3）校正以处理潜在的异方差性。

3. 模型 3：注意力再分配检验（H3）

与模型 2 结构相同，但因变量改为特定类别的注意力份额。本文分别对社会类和娱乐类运行 ITS 回归：

$$\text{Share}_{it} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{Trend}_t + \beta_3 (\text{Post}_t \times \text{Trend}_t) + \gamma_w + \delta_q + \theta \text{Holiday}_t + \varepsilon_t \quad (9)$$

根据 H3a 和 H3b，预期社会类的 $\beta_1 < 0$（份额即时下降），娱乐类的 $\beta_1 > 0$（份额即时上升）。模型同样控制星期固定效应、季度固定效应和节假日效应。

4. 模型 4：注意力密度分化检验（H4）

对社会类和娱乐类分别运行 ITS 回归，因变量为单条话题平均在榜时长：

$$\text{Density}_{it} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{Trend}_t + \beta_3 (\text{Post}_t \times \text{Trend}_t) + \gamma_w + \delta_q + \theta \text{Holiday}_t + \varepsilon_t \quad (10)$$

根据 H4a 和 H4b，预期社会类的 $\beta_1 \geq 0$（“少而精”），娱乐类的 $\beta_1 > 0$（“多且长”）。各模型均控制星期固定效应、季度固定效应和节假日效应，以提高估计精度并排除时间趋势中的周期性干扰。

以上模型设定充分利用了政策冲击的准自然实验特征，通过 ITS 主效应检验和 DID 稳健性检验的结合使用，从整体层面和类别内部两个维度识别算法治理的因果效应，为后续实证分析提供了严谨的计量框架。

四、实证分析

（一）描述性统计

表3报告了核心变量的描述性统计结果。样本涵盖 2024 年 2 月至 2025 年 10 月共 608 天的日度数据，对应 253,227 条话题观测值。就内容质量而言，社会类话题信息差指数均值达 0.423，即超过四成的社会类话题标题呈现信息诱导特征；娱乐类话题该指标均值为 0.318，相对较低。多样性指标方面，Shannon 熵均值为 0.782（标准差 0.145），HHI 均值为 0.248（标准差 0.062），表明热搜榜注意力分布存在较为明显的集中态势。注意力配置方面，社会类占据 34.82% 的份额，娱乐核心类（含明星、游戏、综艺）占 16.47%，其中明星类占 11.86%。注意力密度方面，社会类单条话题平均在榜时长为 8.45 小时，娱乐核心类为 7.23 小时。

（二）内容质量分化效应检验（H1）

假设 H1 预测分级分类治理将导致社会类内容质量提升而娱乐类保持稳定。本文首先采用 ITS 方法检验社会类信息差指数的整体变化（H1a），继而检验娱乐类质量指标的稳定性（H1c）。

表4 Panel A 报告了社会类信息差指数的 ITS 回归结果。列（1）显示，信息差指数在政策后即时下降 0.0847（$p < 0.001$），降幅达 20.0%；趋势项系数为-0.0008（$p = 0.015$），信息诱导现象呈持续改善态势。列（2）显示，机构媒体占比即时上升 0.0672（$p < 0.001$），增幅为 25.2%。Panel B 显示，娱乐类信息差指数的即时效应为-0.0123（$p = 0.187$），趋势效应为-0.0002（$p = 0.624$），均不显著，表明娱乐类内容质量在监管前后未发生实质性变化，支持 H1c。

质量分化效应的形成可从资源依赖视角加以理解。第一，差异化合规成本：社会类内容因其公共属性与舆情风险被纳入更严格的治理序列，平台为规避监管红线而大幅强化审核力度，合规成本显著上升。第二，资源依赖与挤出效应：高风险领域成为平台须优先满足的核心依赖，治理资源向社会类倾斜，对娱乐类形成挤出。第三，信号显示与合法性获取：平台借助社会类质量提升向监管者传递合规信号，而娱乐类因监管压力较小而维持既有质量水平。

表 3 核心变量描述性统计

变量	观测数	均值	标准差	最小值	最大值
质量指标（*H1*）					
InfoGap（社会类）	211,842	0.423	0.494	0	1
InfoGap（娱乐类）	41,667	0.318	0.466	0	1
Institutional 占比（社会类）	608	0.267	0.089	0.103	0.512
多样性指标（*H2*）					
Shannon 熵	608	0.782	0.145	0.412	1.134
HHI	608	0.248	0.062	0.156	0.487
单条话题重复上榜次数	608	2.34	0.67	1.12	4.56
连续在榜时长（小时）	608	6.78	2.13	3.21	14.23
注意力配置指标（*H3*）					
Share（社会类）	608	0.3482	0.0734	0.189	0.523
Share（娱乐核心类）	608	0.1647	0.0456	0.078	0.289
Share（明星类）	608	0.1186	0.0389	0.052	0.234
Share（游戏类）	608	0.0312	0.0145	0.011	0.078
TotalHeat（百万）	87	12.45	2.34	7.89	18.67
注意力密度指标（*H4*）					
社会类总在榜时长（小时/日）	608	82.3	18.7	42.1	134.5
Density（社会类，小时）	608	8.45	2.13	4.56	15.23
娱乐核心类总在榜时长（小时/日）	608	38.6	12.4	16.8	72.3
Density（娱乐核心类，小时）	608	7.23	1.89	3.78	13.45

注：观测数 608 对应日度数据，253,227 对应话题级观测值，87 对应周度数据；信息差指数为二元变量，均值代表信息诱导比例。

表 4 内容质量分化效应（ITS 回归）

Panel A：社会类质量提升（*H1a*）		
因变量	(1) *InfoGap*	(2) *Institutional* 占比
Post（即时效应）	−0.0847***	0.0672***
	(0.0156)	(0.0089)
Trend（趋势）	0.0003	−0.0004
	(0.0003)	(0.0002)
Post × *Trend*（趋势效应）	−0.0008*	0.0006**
	(0.0003)	(0.0002)
常数项	0.423***	0.267***
	(0.0104)	(0.0059)
N	608	608
R^2	0.524	0.618
Panel B：娱乐类质量稳定（*H1c*）		
因变量	(3) *InfoGap*	
Post（即时效应）	−0.0123	
	(0.0093)	
Post × *Trend*（趋势效应）	−0.0002	
	(0.0004)	
N	608	
R^2	0.312	

注：$^{***}p < 0.001$，$^{**}p < 0.01$，$^{*}p < 0.05$；括号内为 Newey-West 标准误。

（三）注意力去中心化效应检验（H2）

假设 H2 预测算法治理将削弱头部话题的注意力垄断，提升热搜榜的多样性水平。本文以 2024 年 11 月 12 日为政策断点，采用 ITS 方法对 Shannon 熵、HHI、单条话题重复上榜次数及连续在榜时长四项指标进行检验，结果见表5。

列（1）显示，Shannon 熵在政策实施后即时上升 0.146（$p < 0.001$），相对于基准均值 0.782 的增幅为 18.7%；趋势项系数为 0.0012（$p = 0.002$），表明熵值在政策后呈持续上升态势。列（2）显示，HHI 即时下降 0.0292（$p < 0.001$），降幅为 11.8%，趋势项系数为-0.0003（$p = 0.042$），集中度持续走低。列（3）与列（4）分别显示，单条话题重复上榜次数下降 0.547 次（降幅 23.4%），连续在榜时长缩短 2.15 小时（降幅 31.7%），均在 0.1% 水平显著。上述结果表明，算法治理通过限制单一话题的重复上榜频次与在榜时长，有效瓦解了头部垄断格局，推动了注意力分布的去中心化。

就机制而言，去中心化主要通过两条路径实现。其一，供给侧约束：《算法专项治理清单指引》明确规定平台“不得利用算法操纵热点话题”，对单一话题垄断榜单形成硬约束。其二，需求侧多样性激励：监管同时要求“保障信息内容多样性”，平台为规避单一话题过度曝光可能招致的监管风险，主动提升长尾话题的推荐权重。两条路径相互叠加，共同推动了注意力分布的均衡化。

表 5 注意力去中心化效应（ITS 回归）

因变量	(1) Shannon 熵	(2) *HHI*	(3) 重复上榜次数	(4) 连续在榜时长
Post（即时效应）	0.146***	−0.0292***	−0.547***	−2.15***
	(0.0234)	(0.0056)	(0.089)	(0.342)
Trend（趋势）	−0.0008	0.0002	−0.0021	0.0156
	(0.0005)	(0.0001)	(0.0018)	(0.0089)
Post × *Trend*（趋势效应）	0.0012**	−0.0003*	−0.0089***	−0.0234**
	(0.0005)	(0.0001)	(0.0024)	(0.0098)
常数项	0.782***	0.248***	2.341***	6.781***
	(0.0156)	(0.0037)	(0.059)	(0.223)
N	608	608	608	608
R^2	0.742	0.681	0.635	0.587
Newey-West 标准误	✓	✓	✓	✓

注：$^{***}p < 0.001$，$^{**}p < 0.01$，$^{*}p < 0.05$；括号内为 Newey-West 标准误；*Post* 为政策后虚拟变量（2024 年 11 月 12 日后取 1）；*Trend* 为时间趋势（政策前归一化为 0）。

（四）注意力再分配效应检验（H3）

假设 H3 预测，在注意力零和约束下，对社会类的监管强化将引发注意力向娱乐类的系统性转移，形成“水床效应”。

表6报告了各类别注意力份额的 ITS 回归结果。列（1）显示，社会类份额即时下降 0.0316（$p < 0.001$），降幅为 9.1%，趋势项系数为-0.0008（$p < 0.001$），份额呈持续走低态势。列（2）显示，娱乐核心类份额即时上升 0.0384（$p < 0.001$），增幅达 23.3%，趋势项系数为 0.0006（$p = 0.003$）。进一步分解表明，明星类份额即时上升 0.0361（$p < 0.001$），游戏类上升 0.0112（$p = 0.002$），两者构成娱乐类扩张的主体。

表 6 注意力再分配效应（类别份额 ITS 回归）

因变量	(1) 社会类	(2) 娱乐核心类	(3) 明星类	(4) 游戏类	(5) 其他类
Post（即时效应）	−0.0316***	0.0384***	0.0361***	0.0112**	−0.0068
	(0.0045)	(0.0038)	(0.0042)	(0.0035)	(0.0051)
Trend（趋势）	0.0002	−0.0001	−0.0001	0.0000	−0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Post × *Trend*	−0.0008***	0.0006**	0.0005**	0.0002*	0.0002
	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0002)
常数项	0.3482***	0.1647***	0.1186***	0.0312***	0.4871***
	(0.0030)	(0.0025)	(0.0028)	(0.0023)	(0.0034)
N	608	608	608	608	608
R^2	0.693	0.728	0.715	0.542	0.398

注：$^{***}p < 0.001$，$^{**}p < 0.01$，$^{*}p < 0.05$；括号内为 Newey-West 标准误。

表7报告了周度总热度的断点检验结果。即时效应为 0.0124（$p = 0.626$），趋势效应为-0.0089（$p = 0.734$），均不显著。这一结果表明，尽管注意力在不同类别间发生了显著再分配，但总量保持稳定，印证了注意力零和约束的存在。监管引发的是注意力的结构性调整而非总量变动。

表 7 周度总热度守恒检验（ITS 回归）

因变量	*TotalHeat*（百万）
Post（即时效应）	0.0124
	(0.0253)
Trend（趋势）	0.0234
	(0.0189)
Post × *Trend*（趋势效应）	−0.0089
	(0.0261)
常数项	12.452***
	(0.167)
N	87
R^2	0.234
p 值（*Post*）	0.626
结论	接受总量守恒假设

注：括号内为 Newey-West 标准误；总热度为所有话题热度值的周度加总。

上述结果验证了“水床效应”机制：监管提升社会类审核成本的同时，多样性约束限制了单一话题的垄断能力；平台在新约束下重新优化注意力配置，将受挤压的社会类配额再分配至审核成本较低的娱乐类，形成“社会类紧缩—娱乐类扩张”的跷跷板格局。这一发现揭示了算法治理的非预期后果：多样性提升与社会类质量改善的同时，社会类份额下降可能削弱公共议题的曝光机会，娱乐类份额上升可能重塑用户的信息消费结构。

（五）注意力密度分化效应检验（H4）

假设 H4 预测，在注意力再分配过程中，社会类呈现“少而精”调整，娱乐类呈现“多且长”调整。

表8报告了注意力密度指标的分组ITS回归结果。Panel A 显示，社会类总在榜时长的即时效应为-12.3 小时/日（$p = 0.210$），不显著；但时间份额即时下降 0.0316（$p < 0.001$），单条话题平均在榜时长即时上升 2.45 小时（$p < 0.001$）。这表明社会类在总量稳定的情况下，通过压缩上榜话题数量、延长单条话题时长，实现了“少而精”的调整路径。Panel B 显示，娱乐核心类总在榜时长即时上升 18.7 小时/日（$p < 0.001$），时间份额上升 0.0384（$p < 0.001$），单条话题平均在榜时长上升 1.89 小时（$p < 0.001$）。娱乐类在总量、份额、密度三个维度均呈显著上升态势，形成“多且长”的调整格局。

密度分化揭示了平台在注意力再分配过程中的微观调整逻辑。社会类的“少而精”源于监管的双重约束：质量门槛淘汰低质内容（分母减少），多样性约束限制单一话题垄断但不禁止高质量话题获得较长时长（分子相对稳定），综合效应是密度上升。娱乐类的“多且长”源于水床效应的注意力溢出：平台通过增加娱乐类话题数量与单条话题时长来吸收从社会类溢出的配额，形成“蓄水池”效应。

（六）稳健性检验

为验证上述结果的可靠性，本文进行了多项稳健性检验。

表 8 注意力密度分化效应（分组 ITS 回归）

Panel A：社会类“少而精”调整			
因变量	(1) 总在榜时长	(2) *Share*	(3) *Density*
Post（即时效应）	−12.3	−0.0316***	2.45***
	(9.78)	(0.0045)	(0.456)
Post × *Trend*	−0.234	−0.0008***	0.0234**
	(0.189)	(0.0002)	(0.0089)
N	608	608	608
R^2	0.412	0.693	0.635
Panel B：娱乐核心类“多且长”调整			
因变量	(4) 总在榜时长	(5) *Share*	(6) *Density*
娱乐核心类			
Post（即时效应）	18.7***	0.0384***	1.89***
	(3.45)	(0.0038)	(0.389)
Post × *Trend*	0.156**	0.0006**	0.0156*
	(0.067)	(0.0002)	(0.0078)
明星类			
Post（即时效应）	16.4***	0.0361***	1.67***
	(2.89)	(0.0042)	(0.345)
游戏类			
Post（即时效应）	5.2**	0.0112**	0.84*
	(1.95)	(0.0035)	(0.389)
N	608	608	608
R^2	0.567	0.728	0.542

注：$^{***}p < 0.001$，$^{**}p < 0.01$，$^{*}p < 0.05$；括号内为 Newey-West 标准误；总在榜时长单位为小时/日。

1. DID 稳健性检验与平行趋势检验（H1b）

表9报告了 H1b 的 DID 回归结果。交互项系数为-0.0205（$p < 0.001$），表明政策后机构媒体相对自媒体的信息差净下降 2.05 个百分点，验证了平台 “源头筛选” 机制的有效性。

表 9 机构媒体相对优势 DID 检验（H1b）

因变量	*InfoGap*
Post × *Institutional*（DID 估计量）	−0.0205***
	(0.0032)
Institutional	−0.0734***
	(0.0089)
Post	−0.0642***
	(0.0123)
话题固定效应	✓
时间固定效应	✓
N	211,842
R^2	0.457

注：***$p < 0.001$；括号内为聚类到话题层面的稳健标准误。

图2报告了平行趋势检验结果。政策实施前（-9 至-1 期），交互项系数围绕零值波动且置信区间均包含零，表明机构媒体与自媒体在政策前不存在显著的差异化趋势，平行趋势假设成立。政策实施后（0 至 10 期），系数显著为负，表明政策效应在实施后即刻显现并持续存在。

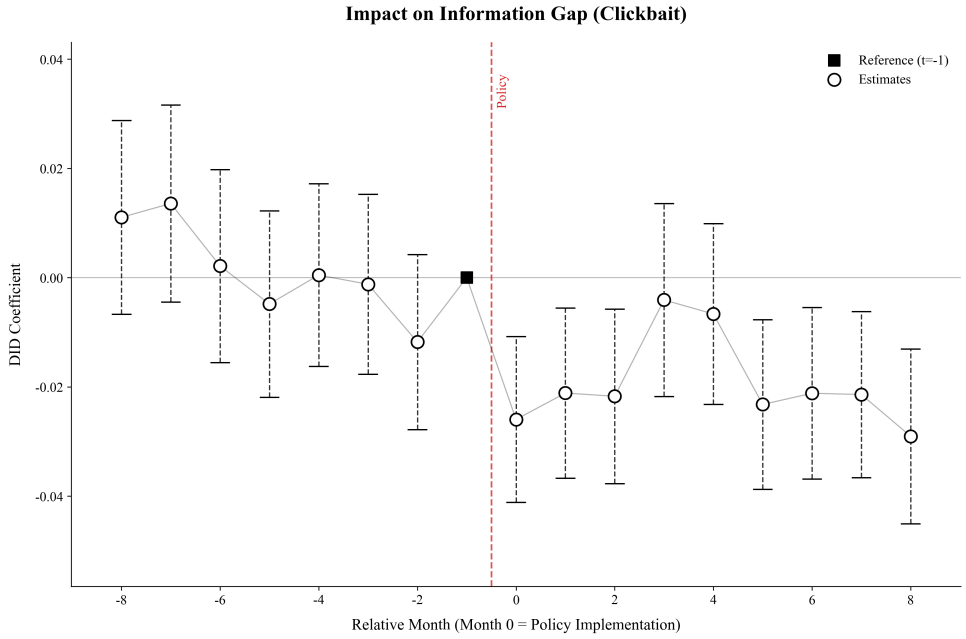

图 2 平行趋势检验：信息差指数的动态 DID 估计

注：图中展示了以政策实施前一期（$t = -1$）为基准的动态 DID 估计系数及 95% 置信区间。空心圆点表示各期 DID 系数估计值，虚线为 95% 置信区间，实心方块表示基准期（$t = -1$）。政策实施前各期系数均不显著异于零，支持平行趋势假设；政策实施后系数显著为负，表明政策效应持续存在。

2. 其他稳健性检验

表10汇总了其他稳健性检验结果。

表 10 稳健性检验汇总

检验类型	(1) 安慰剂	(2) 替代指标	(3) 断点敏感性	(4) 异质稳健 SE
因变量	Shannon 熵	Gini 系数	*Share*（社会类）	各主效应
核心系数	0.0123	-0.0456^{***}	$[-0.0298, -0.0334]$	方向一致
标准误	(0.0167)	(0.0089)	均$p < 0.01$	显著性不变
p 值	0.728	<0.001	全部<0.01	—
结论	通过	通过	通过	通过

注：列（1）伪断点为 2024 年 9 月 1 日；列（2）以 Gini 系数替代 Shannon 熵；列（3）断点在 2024 年 11 月 12 日±7 天窗口内移动；列（4）以异质稳健标准误替代 Newey-West 标准误。

第一，安慰剂检验。以 2024 年 9 月 1 日为伪断点重新估计 ITS 模型，Shannon 熵的即时效应为 0.0123（$p = 0.728$），不显著，表明政策前不存在结构性断点。

第二，替代指标检验。以 Gini 系数替代 Shannon 熵和 HHI，政策后即时下降 0.0456（$p < 0.001$），与主要结论一致。

第三，断点敏感性检验。将断点在 ±7 天窗口内移动，社会类份额的即时效应在 $[-0.0298, -0.0334]$ 区间内，均在 1% 水平显著。

第四，异质稳健标准误。以异质稳健标准误替代 Newey-West 标准误，核心结论均保持稳健。

此外，本文还进行了以下补充检验（结果备索）：排除重大节假日样本、缩短样本期至政策前后各 90 天、以 Poisson 回归替代 OLS 估计计数型因变量，核心结论均保持稳健。

（七）小结

综上，实证分析全面验证了四项核心假设。H1 检验表明分级分类治理导致社会类质量显著提升而娱乐类保持稳定，DID 估计和平行趋势检验进一步强化了因果识别；H2 检验表明算法治理有效推动了注意力分布的去中心化（Shannon 熵上升 18.7%，HHI 下降 11.8%）；H3 检验表明在注意力零和约束下社会类份额下降 3.16 个百分点、娱乐类上升 3.84 个百分点，总热度守恒印证了“水床效应”；H4 检验表明社会类呈“少而精”调整、娱乐类呈“多且长”调整。这些发现共同刻画了算法治理对平台生态系统的多维重构效应，验证了基于资源依赖理论的分析框架。

五、异质性分析

为进一步探讨算法治理效应的异质性，本文从季节周期、时间段、话题热度三个维度展开分析，结果见表11。

（一）季度异质性

信息差指数的政策效应呈现显著的季度差异。Q2（4-6 月）效应最为强烈，即时下降 2.45 个百分点（$p < 0.001$）；Q1（1-3 月）次之，即时下降 1.29 个百分点（$p = 0.050$）；

表 11 异质性检验结果（ITS 回归）

异质性维度	分组	即时效应	p 值	趋势效应	p 值
Panel A：季度异质性（信息差指数）					
	Q1	-0.0129^{*}	0.050	0.0001	0.587
	Q2	-0.0245^{***}	<0.001	-0.0001	0.491
	Q3	-0.0021	0.641	-0.0000	0.566
	Q4	-0.0087	0.160	-0.0001	0.657
Panel B：时间段异质性（情绪唤起指数）					
	工作日	0.0013	0.725	-0.0000	0.915
	周末	-0.0129^{*}	0.017	-0.0000	0.711
Panel C：话题热度异质性（信息差指数）					
	高热度	-0.0074	0.196	-0.0000	0.461
	中热度	0.0056	0.290	0.0001^{**}	0.006
	低热度	0.0055	0.304	0.0000	0.167

注：$^{***}p < 0.001$，$^{**}p < 0.01$，$^{*}p < 0.05$；样本为社会类话题；所有模型均控制季度固定效应与节假日效应；热度分组基于三分位数划分。

Q3 与 Q4 效应则不显著。这一季节性模式可能与监管资源配置周期有关：Q1-Q2 通常是监管部门年度考核与专项行动密集期，平台合规压力较大；下半年监管力度相对趋缓，治理效应有所减弱。

（二）时间段异质性

情绪唤起指数在工作日与周末呈现差异化响应。周末情绪唤起指数即时下降 1.29 个百分点（$p = 0.017$），而工作日效应不显著（$p = 0.725$）。这一发现揭示了用户行为与算法推荐的交互模式：周末用户在线时长较长、信息消费更为集中，平台在此时段更倾向于推送高情绪价值内容以提升用户黏性；政策实施后，周末的情绪激励策略受到更大约束，效应更为显著。

（三）话题热度异质性

按热度三分位划分样本后，高热度话题呈现即时下降趋势但不显著（$p = 0.196$）；中热度话题在趋势效应上显著为正（$p = 0.006$），表明政策后中热度话题的信息差呈持续上升态势；低热度话题两项效应均不显著。这一非对称格局暗示，算法治理主要约束了头部高热度话题的信息诱导策略，而中热度话题作为“腰部内容”在监管盲区中获得了更大的策略空间。

六、结论与讨论

本文利用 2024 年 11 月中国清朗行动算法治理政策这一准自然实验，构建基于资源依赖理论的统一分析框架，采用断点回归时间序列（ITS）与双重差分（DID）相结合的识别策略，基于微博热搜榜 608 天、超过 253,227 条观测值的高频数据，系统评估了算法治理对平台内容生态的多维重构效应。

（一）主要发现

本研究的实证分析全面验证了四项核心假设，揭示了算法治理对平台生态系统的复杂影响：

算法治理在改善内容生态的同时引发了注意力的结构性再分配。具体而言：（1）内容质量分化：分级分类治理导致社会类内容质量显著提升（信息差指数下降 20.0%、机构媒体占比上升 25.2%），DID 估计显示机构媒体相对自媒体的信息差净下降 2.05 个百分点，平行趋势检验支持因果识别的有效性，而娱乐类质量保持稳定，形成差异化的质量改善格局。（2）注意力去中心化：算法治理有效打破了头部话题的垄断格局（Shannon 熵上升 18.7%、HHI 下降 11.8%，单条话题重复上榜次数下降 23.4%、连续在榜时长缩短 31.7%）。（3）水床效应：在注意力总量守恒的零和约束下（周度总热度守恒，$p = 0.626$），社会类份额即时下降 3.16 个百分点、娱乐类即时上升 3.84 个百分点，呈现出典型的“社会类紧缩—娱乐类扩张”跷跷板格局。（4）密度分化：社会类呈“少而精”调整（话题数量减少、单条时长增加），娱乐类呈“多且长”调整（数量和时长同步扩张），揭示了平台在宏观再配置基础上的微观精细化响应策略。

上述发现共同验证了基于资源依赖理论的分析框架：算法治理通过差异化地改变不同内容类型的资源获取成本，触发平台作为资源依赖者的再配置行为，最终导致内容生态的多维重构。多样性提升与质量改善的积极效果，与注意力再分配的非预期后果并存，揭示了算法治理中多元目标之间的潜在张力。

（二）理论贡献

本研究在以下三个方面推进了算法治理与平台经济的理论发展：

1. 构建了基于资源依赖理论的统一分析框架

本研究的核心理论贡献在于将资源依赖理论应用于算法治理情境，构建了统一的分析框架。该框架的核心命题是：用户注意力构成数字平台最核心的稀缺资源，算法治理通过差异化地改变不同内容类型的资源获取成本，触发平台的资源再配置行为。

这一框架的理论价值在于：它将看似分散的多元治理效应——内容质量分化、注意力去中心化、水床效应、密度分化——纳入统一的资源依赖逻辑。四项效应不再是孤立的政策后果，而是平台在新的资源约束下进行系统性再配置的不同维度表现。H1（内容质量分化）是成本重构的直接效应；H2（注意力去中心化）是多样性约束的效应；H3（水床效应）是资源总量约束下的再配置效应；H4（密度分化）是资源再配置的微观表现。

2. 揭示了算法治理的“水床效应”机制

本研究首次在实证层面揭示了算法治理的“水床效应”：在注意力总量约束下，对某一内容领域的监管强化将引发注意力向其他领域的系统性转移。这一发现拓展了平台治理的理论边界，表明监管效果的评估须超越单一维度，关注治理的系统性影响和溢出效应。

水床效应的理论意义在于：它揭示了注意力经济中“零和博弈”的深层逻辑——用户注意力作为稀缺资源，其总量在短期内近似固定，监管改变的是注意力的配置结构而非总量。这一机制为理解算法治理的间接传导路径提供了分析工具：监管并非直接作用

于内容本身，而是通过改变平台的资源获取成本来间接重塑内容生态。

3. 提出了多元治理目标间的权衡框架

本研究基于实证发现，提出算法治理存在多元目标间的潜在权衡。在注意力总量相对稳定的约束下，监管者难以同时最大化以下三个目标：（1）信息多样性：打破头部垄断，提升内容分布的均匀程度；（2）内容质量：提高高风险内容的质量门槛，减少低质内容；（3）公共议题曝光：确保社会类内容获得充分的用户触达。

本研究的实证结果显示，政策在多样性和社会类质量两个维度取得了积极效果，但社会类份额的下降可能削弱公共议题的曝光机会。这一权衡框架为理解算法治理的复杂性提供了理论工具，也为政策设计提供了分析基础。

（三）政策启示

基于上述理论发现，本研究为算法治理的政策优化提出以下建议：

1. 建立多维度治理效果评估体系

当前监管评估多聚焦于单一维度（如内容质量或多样性），本研究表明算法治理具有系统性影响，须建立涵盖多样性、质量、注意力配置、用户效用等多维度的综合评估体系。具体建议包括：（1）将“注意力配置结构”纳入监管评估指标，监测不同内容类型的份额变化；（2）建立“非预期后果”预警机制，识别治理措施可能引发的溢出效应；（3）开展定期的治理效果审计，动态调整监管策略。

2. 探索差异化的正向激励机制

当前监管主要通过提高违规成本来约束平台行为，本研究发现这一“负向激励”模式可能导致水床效应。建议探索正向激励机制作为补充：（1）对高质量社会类内容给予流量扶持或经济激励，弥补监管导致的份额下降；（2）将“公共价值贡献”纳入平台评价体系，激励平台主动提升公共议题的曝光质量；（3）建立“质量—曝光”联动机制，使高质量内容获得与其公共价值相匹配的注意力配额。

3. 增强用户对算法的知情权与选择权

本研究揭示的水床效应本质上是平台在监管约束下的最优化响应，用户在此过程中处于被动地位。建议监管政策在约束平台算法的同时，推动平台向用户提供算法偏好的调整选项：（1）允许用户在一定范围内调整不同内容类型的推荐权重；（2）提供“探索模式”与“偏好模式”的切换功能；（3）增强算法推荐的透明度，让用户了解内容配置的变化及其原因。这种“用户赋权”策略既保障了平台层面的基本多样性，又尊重了用户偏好的异质性。

4. 建立跨平台协同治理机制

本研究聚焦单一平台，但用户的跨平台迁移可能削弱单一平台监管的效果。如果用户因监管导致的内容变化而转向其他平台，则单一平台的治理效果将被稀释。建议在主要平台间建立协同治理机制：（1）统一的内容质量标准和多样性要求；（2）低质内容信息共享和联合治理；（3）同步执法以降低跨平台套利空间。

（四）研究局限与未来方向

本研究存在以下局限，为未来研究提供了方向：

数据局限。第一，用户偏好数据缺失。本研究仅观测到平台供给侧的内容配置变化，无法直接测度用户需求侧的真实偏好变化，未来研究可通过用户调查或行为实验弥补这一不足。第二，跨平台数据缺乏。本研究聚焦微博单一平台，无法验证用户是否因监管而将注意力转移至其他平台，未来研究可收集多平台数据进行比较分析。第三，观测窗口有限。本研究观测窗口为政策实施后约 10 个月，仅能捕捉短期和中期效应，长期效应有待追踪。

未来研究方向。第一，开展用户层面的田野实验，向部分用户提供算法控制权，观测用户的真实选择行为和效用变化，验证“用户赋权”策略的有效性。第二，收集多平台用户行为数据，检验单一平台监管是否导致用户跨平台迁移，评估协同治理的必要性。第三，追踪监管后用户的信息获取行为、政治参与度、主观福利等指标，评估算法治理对用户长期效用的影响。第四，开展国际比较研究，对比中国、欧盟《数字服务法》、美国等不同治理模式的效果差异，识别有效治理实践的共同特征。

（五）结语

算法推荐系统作为数字时代的信息配置机制，深刻影响着公共信息传播和社会认知形成。本研究利用中国清朗行动这一准自然实验，系统评估了算法治理的多维效应，揭示了治理在改善内容生态的同时引发注意力结构性再分配的复杂图景。

“水床效应”的发现具有重要的政策含义：算法治理是一项需要在多元目标间寻求平衡的系统工程，单一维度的优化可能引发非预期后果。在约束平台算法的同时，建立多维评估体系、探索正向激励机制、增强用户知情权与选择权、推进跨平台协同治理，可能是实现更优治理效果的可行路径。

本研究为理解算法治理的复杂效应提供了新的实证证据和理论框架，也为完善平台治理体系提供了政策参考。随着算法技术的持续演进和监管实践的不断深化，算法治理研究将继续面临新的挑战和机遇。

参考文献

Ahmmad, M., Shahzad, K., Iqbal, A., and Latif, M. (2025). Trap of social media algorithms: A systematic review of research on filter bubbles, echo chambers, and their impact on youth. *Societies*, 15(11):301.

Anglin, K. L., Wong, V. C., Wing, C., Miller-Bains, K., and McConeghy, K. (2023). The validity of causal claims with repeated measures designs: A within-study comparison evaluation of differences-in-differences and the comparative interrupted time series. *Evaluation Review*, 47(5):895–931.

Armstrong, M. (2006). Competition in two-sided markets. *The RAND Journal of Economics*, 37(3):668–691.

Berman, R. and Katona, Z. (2020). Curation algorithms and filter bubbles in social networks. *Marketing Science*, 39(2):296–316.

Bucher, T. (2018). *If...Then: Algorithmic Power and Politics*. Oxford University Press, New York.

Chakraborty, A., Paranjape, B., Kakarla, S., and Ganguly, N. (2016). Stop clickbait: Detecting and preventing clickbaits in online news media. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 9–16. IEEE.

Chen, Y., Conroy, N. J., and Rubin, V. L. (2015). Misleading online content: Recognizing clickbait as "false news". In *Proceedings of the 2015 ACM Workshop on Multimodal Deception Detection*, pages 15–19. ACM.

Cima, L., Tessa, B., Trujillo, A., Cresci, S., and Avvenuti, M. (2025). Investigating the heterogeneous effects of a massive content moderation intervention via difference-in-differences. *Online Social Networks and Media*, 48:100320.

Creemers, R. (2022). China's emerging data protection framework. *Journal of Cybersecurity*, 8(1):tyac011.

Cui, H. and Kertész, J. (2023). Competition for popularity and interventions on a Chinese microblogging site. *PLOS ONE*, 18(5):e0286093.

Davenport, T. H. and Beck, J. C. (2001). The attention economy. *Ubiquity*, 2001(May):1–es.

Degen, K. and Gleiss, A. (2025). Time to break up? the case for tailor-made digital platform regulation based on platform-governance standard types. *Electronic Markets*, 35(1):5.

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2):56–62.

Ghazawneh, A. and Henfridsson, O. (2013). Balancing platform control and external contribution in third-party development: The boundary resources model. *Information Systems Journal*, 23(2):173–192.

Gillespie, T. (2014). The relevance of algorithms. In Gillespie, T., Boczkowski, P. J., and Foot, K. A., editors, *Media Technologies: Essays on Communication, Materiality, and Society*, pages 167–194. MIT Press, Cambridge, MA.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press, New Haven, CT.

Golman, R. and Loewenstein, G. (2018). Information gaps: A theory of preferences regarding the presence and absence of information. *Decision*, 5(3):143–164.

Gomes, A. B. and Sultan, A. (2024). Problematizing content moderation by social media platforms and its impact on digital harm reduction. *Harm Reduction Journal*, 21(1):194.

Gorwa, R., Binns, R., and Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1):1–15.

Gritsenko, D. and Wood, M. (2022). Algorithmic governance: A modes of governance approach. *Regulation & Governance*, 16(1):45–62.

Guess, A. M., Malhotra, N., Pan, J., Barberà, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., et al. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, 381(6656):398–404.

He, S., Hollenbeck, B., and Proserpio, D. (2022). The market for fake reviews. *Marketing Science*, 41(5):896–921.

Heitmayer, M. (2025). The second wave of attention economics: Attention as a universal symbolic currency on social media and beyond. *Interacting with Computers*, 37(1):18–29.

Hindman, M. (2018). *The Internet Trap: How the Digital Economy Builds Monopolies and Undermines Democracy*. Princeton University Press, Princeton, NJ.

Hosanagar, K., Fleder, D., Lee, D., and Buja, A. (2014). Will the global village fracture into tribes? recommender systems and their effects on consumer fragmentation. *Management Science*, 60(4):805–823.

Hosseinmardi, H., Ghasemian, A., Rivera-Lanas, M., Horta Ribeiro, M., West, R., and Watts, D. J. (2024). Causally estimating the effect of YouTube's recommender system using counterfactual bots. *Proceedings of the National Academy of Sciences*, 121(8):e2313377121.

Huber, T. L., Fischer, T. A., Dibbern, J., and Hirschheim, R. (2013). A process model of complementarity and substitution of contractual and relational governance in IS outsourcing. *Journal of Management Information Systems*, 30(3):81–114.

Jacobides, M. G., Cennamo, C., and Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal*, 39(8):2255–2276.

Jhaver, S., Bruckman, A., and Gilbert, E. (2019). Does transparency in moderation really matter? user behavior after content removal explanations on Reddit. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–27.

Jiang, H., Rehm, J., Tran, A., and Lange, S. (2024). Interrupted time series design and analyses in health policy assessment. *MedRxiv*. Preprint.

Kango, U. (2025). Algorithmic governance. In *Handbook of Human-Centered Artificial Intelligence*, pages 1–26. Springer Nature Singapore, Singapore.

Katyal, S. K. (2019). Private accountability in the age of artificial intelligence. *UCLA Law Review*, 66:54–141.

Leskovec, J., Backstrom, L., and Kleinberg, J. (2009). Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 497–506. ACM.

Liu, X., Zhao, J., Li, Z., Wang, D., and Chen, A. (2025). Unveiling the making of trending topics on a digital platform: A research note on Chinese Sina Weibo. *Social Science Computer Review*.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, 116(1):75–98.

Mousavi, S., Gummadi, K. P., and Zannettou, S. (2024). Auditing algorithmic explanations of social media feeds: A case study of TikTok video explanations. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, volume 18, pages 1110–1122.

Munger, K. (2020). All the news that's fit to click: The economics of clickbait media. *Political Communication*, 37(3):376–397.

Oestreicher-Singer, G. and Sundararajan, A. (2012). The visible hand? demand effects of recommendation networks in electronic markets. *Management Science*, 58(11):1963–1981.

O'Reilly, T., Strauss, I., and Mazzucato, M. (2024). Algorithmic attention rents: A theory of digital platform market power. *Data & Policy*, 6:e6.

Panigutti, C., Yela, D. F., Porcaro, L., Bertrand, A., and Garrido, J. S. (2025). How to investigate algorithmic-driven risks in online platforms and search engines? a narrative review through the lens of the EU Digital Services Act. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*, pages 828–839. ACM.

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin UK.

Parker, G. G. and Van Alstyne, M. W. (2018). Innovation, openness, and platform control. *Management Science*, 64(7):3015–3032.

Pfeffer, J. and Salancik, G. R. (1978). *The External Control of Organizations: A Resource Dependence Perspective*. Harper & Row, New York.

Rochet, J.-C. and Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4):990–1029.

Sandvig, C., Hamilton, K., Karahalios, K., and Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. In *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*. International Communication Association.

Sheehan, M. (2023). China's AI regulations and how they get made. *Horizons: Journal of International Relations and Sustainable Development*, (24):108–125.

Simon, H. A. (1996). Designing organizations for an information-rich world. *International Library of Critical Writings in Economics*, 70:187–202.

Sunstein, C. R. (2018). *Republic: Divided Democracy in the Age of Social Media*. Princeton University Press, Princeton, NJ.

Tiwana, A., Konsynski, B., and Bush, A. A. (2010). Research commentary—platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information Systems Research*, 21(4):675–687.

Ulbricht, L. and Yeung, K. (2022). Algorithmic regulation: A maturing concept for investigating regulation of and through algorithms. *Regulation & Governance*, 16(1):3–22.

Wang, S., Huang, S., Zhou, A., and Metaxa, D. (2024). Lower quantity, higher quality: Auditing news content and user perceptions on Twitter/X algorithmic versus chronological timelines. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2):1–25.

Webster, J. G. (2016). *The Marketplace of Attention: How Audiences Take Shape in a Digital Age*. MIT Press.

Wu, T. (2017). *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. Vintage Books, New York.

Xu, J. (2024). Opening the 'black box' of algorithms: Regulation of algorithms in China. *Communication Research and Practice*, 10(3):288–296.

Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, 12(4):505–523.

Yu, S. and Sekiguchi, T. (2024). Platform-dependent entrepreneurship: A systematic review. *Administrative Sciences*, 14(12):326.

喻国明 and 刘彧晗 (2023). 从信息竞争到认知竞争：策略性传播范式全新转型——基于元传播视角的研究. 现代传播（中国传媒大学学报）, 45(2):128–134.

喻国明 and 刘彧晗 (2024). 个性化推荐 ≠ 信息茧房：对算法与茧房效应的误读澄清. 青年记者, (7):5–10.

师文 and 陈昌凤 (2023). 平台算法的" 主流化" 偏向与" 个性化" 特质研究——基于计算实验的算法审计. 新闻记者, (11):14–28.

李三希, 张仲元, and 焦倩 (2023). 竞争会降低企业收集消费者信息并增加隐私保护投资吗？. 管理世界, (7):130–147.

焦豪 (2023). 数字平台生态观：数字经济时代的管理理论新视角. 中国工业经济, (7):135–154.